DOCUMENT RESUME

ED 069 754 TM 002 209

AUTHOR	Park, James
TITLE	The Use and Development of Videotape Tests to Assess Achievement of Educational Objectives: Application in Educational Psychology.
PUB DATE	72
NOTE	16p.; Paper presented to the CAPE Conference, Montreal, 1972
AVAILABLE FROM	Information Section, Schools Council, 160 Great Portland Street, London WIN 6LL, England (free)
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Academic Achievement; Audiovisual Aids; College Students; Educational Psychology; Educational Testing; *Empathy; *Group Dynamics; Interaction Process Analysis; *Measurement Instruments; Measurement Techniques; Multiple Choice Tests; Questionnaires; Speeches; Statistical Analysis; Technical Reports; Test Construction; *Testing; Test Reliability; Tests; *Video Tape Recordings; Visual Measures
IDENTIFIERS	Corkhuffs Empathy Scale; Saint Sebastian Syndrome

ABSTRACT

The use of videotape tests is presented. Such tests enable the educator to assess student performance more directly than traditional paper and pencil tests. Test 1 was exploratory. Test 2 was designed to measure empathetic understanding. It contains 16 scenes, each about one minute long, which show five individuals in a group situation. The subject taking the test considers himself the 6th member of the group and responds at the end of each scene (1) to record responses which show a high degree of communication of empathetic understanding, and (2) to select from five alternatives the response which shows the highest degree of empathetic understanding. Results of the free response version showed an inter-rater reliability of .95. Correlation of the multiple choice version with the Carkhuff Empathy Scale was modest, .56. Test 3 attempted to assess understanding of group dynamics. It demonstrates that some measurement of observational understanding is possible, but is still in the experimental stage. Tests 4 and 5 are experiments in videotape segments used to determine achievement in educational psychology. Although they are not developed enough to report reliability, responses to student questionnaires regarding them indicate the testing method is useful. (DJ)

ED 069754

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

170

SCOPE OF INTEREST NOTICE
The ERIC Facility has assigned this document for processing to: TM EM
In our judgement, this document is also of interest to the clearinghouses noted to the right. Indexing should reflect their special points of view.

THE USE AND DEVELOPMENT OF VIDEOTAPE TESTS TO ASSESS ACHIEVEMENT OF EDUCATIONAL OBJECTIVES: APPLICATION IN EDUCATIONAL PSYCHOLOGY

by

James Park

McGill University

TM 002 209

A paper presented to the C.A.P.E. Conference, Montreal, 1972

COPIES OF THIS DOCUMENT ARE AVAILABLE GRATIS FROM INFORMATION SECTION, SCHOOLS COUNCIL, 160 GREAT PORTLAND STREET, LONDON W1N 6LL, ENGLAND.

Over the past few years an increasing number of educators have discontinued the practice of using classroom tests or major final examinations to assess the impact of their instruction. The effects on pupil performance that are associated with this trend are somewhat mixed. The most beneficial result is a more relaxed climate for learning wherein students work, often enthusiastically, at various projects which demonstrate some increase in their competence. The anxiety of having to develop a meaningful project in line with the student's interest is considerably less than that associated with threatening examinations. On the other hand, a minority of students take advantage of the project or term paper oriented assessment. They submit projects which have been done by peers or professionals elsewhere (Term paper banks in fraternity houses and agencies which sell "guaranteed pass" papers are not unknown). There is also some reason to suspect that when students are not tested for subject mastery, the qualities which make them human contribute to the avoidance of energy expenditure on the integration of certain concepts. The exam provides a form of extrinsic motivation for the student to overcome natural tendencies to follow the Law of EFFECT!

There are still two strong arguments for the use of examinations at the university level. Firstly, the university has a responsibility to the public in guaranteeing that graduates can indeed perform skills which they claim to be competent in doing. Secondly, university instructors have a responsibility to their students to determine what impact, if any, their particular method of instruction has on cognitive change. While realizing the fact that certain pupil qualities such as intelligence,

initiative, and so on account for the major portion of variance in achievement, there is still a need to view the instructor and the methods he employs as independent variables.

In spite of considerable pressure from students to do away with examinations, it becomes the task of the educator to set examination situations which are acceptable to students and which allow for the assessment of achievement of educational objectives. This task is difficult, but not necessarily impossible. This paper discusses the use of videotape tests as an alternative approach from the more traditional paper and pencil test. The need to develop other approaches should also be apparent. While the paper is specifically concerned with the development and use of these tests in various Educational Psychology courses, it is implied that similar use of videotape procedures for testing purposes can and should be made in other subject areas.

One argument against the use of tests in Education courses which has been advanced loudly by students is that performance on examinations provides little indication of classroom performance. This argument has merit. A gap does indeed exist between the university tower and the urban classroom. The realities of the two situations are different. These differences are often reflected in examinations. For instance, the child described in written case material presented on examinations for student analysis may bear little resemblance to the "live" child.

While educators have slowly come to appreciate the advantages and disadvantages of audio-visual aids <u>for instruction</u>, the literature suggests that little use of that equipment has been made <u>for the assessment</u> of instruction. Exams characteristically lack input of visual and auditory cues.

This paper supports the contention that examinations can be advanced two steps closer to "reality" by paying more careful attention to the increase of such sensory inputs.

My own experience in the development of videotape tests is quite limited, but worth sharing. Over the past two years I have been involved with the development of five videotape tests. Three of the tests were rather sophisticated productions which were undertaken with staff and students of the University of Alberta. They were designed to measure attainment of educational objectives in various sections of special educational psychology courses. The intended learning outcomes of these courses involved attempts to increase student competence in understanding group dynamics and the communication of empathy. More recently, I hastily developed two short videotape tests for use in a more traditional senior level Educational Psychology course at McGill. These tests and the procedures employed in their development are described briefly below as follows:

TEST 1

Our first attempt at the construction of a videotape test was in June of 1970. Drs. Eberlein, Matheson, several graduate students and myself had heard of work in California where videotape tests were being employed to assess counsellor training. In those tests, counsellor trainees were asked to respond to a one minute segment of a taped client. After one minute the trainee had to make a verbal response to the taped client.

As a variation of this theme our first videotape test consisted of twenty or so scenes wherein groups of five adults were roleplaying a group problem. For example, one scene portrayed problems of handling dominant

individuals in groups. Another scene was concerned with the St. Sebastian" syndrome wherein one member controls the group's airtime by answering all questions in an ambiguous manner. At the end of each scene subjects taking the test were required to make a "helpful" comment to the group. Our original intention was to use Carkhuff's various scoring scales to assess how facilitative each comment was. Upon administering that test in a pilot-study, it became apparent that this first test was inadequate in certain respects. The instructions were not specific enough, the interval between scenes was too short, too many members on the tape spoke, and so on. With the benefit of knowledge acquired in producing that videotape, Wayne, Matheson and myself produced a second videotape which was designed to measure increases in empathetic understanding (using Carkhuff's Empathy Scale for rating purposes).

TEST 2 - The Park-Matheson Human Relations Videotape Test of Empathetic Understanding (HRVT)

In brief, this test contains 16 scenes, each of which shows 5 individuals in a group situation. Twelve scenes are from the second videotape, four are from the first test. The viewer is instructed to consider himself as the sixth member of the group. Each scene lasts approximately one minute and shows one or more of the group members expressing their personal feelings about some problem or situation. At the end of each scene the screen goes blank for one minute while the subject taking the test responds to a designated group member. The test has been developed so as to allow the respondent to follow two kinds of instruction:

1. The viewer is asked to write responses which show a high degree of communication of empathic understanding;

subsequently,

2. The viewer is asked to select from five alternatives the response which shows the highest degree of empathic understanding.

In our experimental course students were first required to complete the free response version. They were then re-shown the videotape and completed the multiple choice version. This procedure was followed at the beginning and end of the course.

A. The Free Response Version of the HRVT's Ability to Express Empathic Understanding

To determine whether or not any of the students had indeed increased in their ability to communicate empathic understanding, the written responses obtained from the pre- and post-testing sessions were typed individually on 5 x 8 sheets. This procedure was adopted so as to reduce the possibility of raters being influenced by handwriting differences. Two trained raters undertook a blind analysis of the data. Initial correlation of their ratings for 2908 responses was .70. After discussion - and re-analysis of 220 responses on which the raters had differed by more than .5 points (on a 5 point scale), the inter-rater reliability was increased to .95. The split-half reliabilities were .88 and .92 for the pre and post-tests, respectively, without applying a Spearman-Brown correction for length.

B. The Multiple-Choice Version of the HRVT: Ability to Recognize Empathic Understanding

For the multiple choice version of the HRVT, students were asked to select from five alternatives the response which shows the highest level

of empathy to a designated group member in each scene. The alternatives for each scene were selected from the responses of various individuals who volunteered to preview the tape. These individuals included professional counsellors, professors, graduate and undergraduate students. An inspection of their responses to the test indicated that items showing various levels of empathy were available for each scene. Due to time considerations only 10 of the 16 scenes were used in the testing prior to the course; however, the post-test was lengthened to 16 items.

In its current stage of development the multiple choice version of the HRVT is a "best" answer test which is scored on an "all or nothing" basis. That is to say, subjects receive "1" if they are able to choose the response which, according to the "expert" group of assessors, shows the highest degree of empathic understanding to the designated group member on videotape. The choice of any other alternative receives a "0" mark. The maximum possible score was 10 on the pre-test and 16 on the post-test.

The pre-test and post-test means were 3.6 and 6.7, respectively. Scores ranged between 0 and 7 on the pre-test, and 0 and 11 on the post-test. The K-R_{20} reliability for the pre-test was .54. This statistic increased slightly to .58 probably due to lengthening of the test.

The results obtained on the HRVT were then compared with results obtained on Carkhuff's paper and pencil discrimination test which supposedly measures attainment of similar skills. A moderate Pearson product moment correlation of .42 was observed between the two tests for the same students (N=98). The product moment correlations were .56 between the two versions of the HRVT given at the end of the course. A higher correlation between the two forms of the HRVT was not anticipated, as the free response

purports to measure ability to express empathy in contrast to the multiple choice version which measures ability to recognize expressed empathy.

TEST 3 - The Assessment of Understanding of Group Dynamics

One of the prime goals of human relations training courses is to improve understanding of the processes clustered loosely under the title "group dynamics". While advocates of training programs claim that their treatments increase one's sensitivity to group processes, there appears to be only limited or no evidence that such claims are valid. Except for scattered attempts to develop an increased understanding of group-decision making strategies using as "treatments" such instruments as "The Twelve Angry Men" film, there has been little or no attempt to construct instruments designed to assess the cognitive understanding of behavior in groups, prior to the work described here.

During the pilot run which preceded our experimental course, members of the research team expressed a yearning for the day when someone would develop a sensible approach to assessing understanding of group process. A rather simple solution to this difficult assessment problem was proposed during a "brain-storming" session. It was hypothesized that if group members really do acquire cognitive understanding during human relations training, they should be able to demonstrate this understanding by correctly categorizing the ongoing behavior of a similar group. Starting from this premise, Drs. McLeish, Matheson and myself decided that it should be possible for "experts" (in group dynamics) to view the videotaped interaction of a group and to reach a consensus about prevailing behavior patterns and the ongoing dynamics. The extent to which relatively more naive subjects, viewing such a videotape, choose

from a number of alternatives the same description as the experts would be indicative of their cognitive understanding of group processes.

To implement this idea, two self-analytic training groups in a summer pilot project were videotaped. Each group was videotaped for one hour, the students were summer school undergraduate teachers-in-training. Of the videotapes, one in particular seemed quite rich in displaying various interaction themes, including what would be termed scapegoating, fantasy, projected aggression, and so on. This tape was therefore chosen for further development as a group process test.

A preliminary analysis was made by combing through the tape six or seven times, looking for what might be considered natural or logical break-points. These were points in the group where different themes appeared. Having provisionally decided on these, several professors and doctoral students were asked to assist in providing interpretations of the scenes isolated between the defined logical break-points. Most of this group had extensive experience in human relations training groups and/or therapy groups; they professed to represent several schools of thought about group processes. The videotape was played to this group and stopped at the various breaking-points. Each member of the group was asked to write a short description of each segment; these were then discussed. A remarkable amount of agreement was expressed in the discussions after each segment. Bearing the suggestions of this group in mind, it was possible to identify eleven distinct segments and to develop seventeen four-item multiple choice questions. This test was called the Group Process Analysis Test (GPAT).

After deliberation and a preliminary trial, it was decided that the

GPAT would be quite difficult for most undergraduate students. The answers to the questions depended upon keen detection of verbal clues, sometimes quite minute, provided in the group interaction dialogue. In addition, it was thought possible that some subjects might spend a good deal of time viewing the segments in an overwhelmed condition, possibly even in a state of trepidation. As we were concerned with obtaining a valid assessment of group understanding, it was thought that, ideally the test should be administered twice. Time limitations made it impossible to do this in one session, the GPAT taking 40 minutes to administer. To bypass this difficulty, a transcript of the verbal content of the video-tape discussion was given to each subject during the test session. This transcript was to be used for two purposes: (I) to assist subjects to pick up verbal clues and cues which might normally be missed owing to distracting noises in the classroom; (2) after viewing the videotape and doing the test in the classroom, each subject was to retake the GPAT at home, alone, using the transcript.

Alas, while some things are excellent in theory, reality decreases their value. The first results obtained from the GPAT using a "I", "0", scoring procedure were most discouraging. The Kuder-Richardson 20 post test reliability for the video-tape test taken in class was .19. The KR 20 reliability for answers to the transcript was .52. Further attempts to improve the test were made by trying various other scoring schemes. After considerable thought, a twelve point scale was devised which takes into account the experts', the classes', and the highest 18 students', ratings of each item. Using that scoring system it was possible to sum the students' in-class and at-home scores, and obtain a test with Spearman-Brown split-half reliability of .61.

The GPAT is still considered to be a very rough test of understanding of group dynamics processes. Its value is not in the precision it provides, but rather the concept it presents. It has allowed us to demonstrate that measurement of observational understanding is possible, to some minimal degree at least.

TESTS 4 and 5

The last two tests which I want to discuss are much less sophisticated than these earlier efforts. In fact, they are short videotape segments which I have employed to determine level of achievement in my present Educational Psychology courses. Test 4 is a twelve minute videotape of four senior education students discussing information about two fictitious school students who supposedly applied for a scholarship. Prior to the videotape session I presented four different pieces of information on the students under discussion to the "scholarship committee". The committee however did not realize that they had received information which had different slants. In the videotape session the committee was asked to discuss the problem of awarding a scholarship and decide which of the two students should receive the scholarship. To determine whether or not my present classes had any understanding of various group processes I played the videotape three times to them and asked them to answer ten questions.

The last videotape segment test show a psychologist (myself) interviewing four children from the same family. Each child was interviewed separately and asked various questions. Students taking this videotape test were required to discuss differences observed between the four children with regard to learning, intellectual development, and various signs of maturation. The videotape also lasts approximately

12 minutes. For final examination purposes it was shown twice to the examinees.

I am still in the process of analyzing the results obtained from these two tests and therefore no evidence can be presented here with regard to reliability. There is some evidence however that students, in general, see value in this method of testing. To determine some indication of their feelings towards these last two tests I had the classes respond anonymously on a short questionnaire.

To the question "Do you feel that the videotape questions used here will provide as good an assessment of your ability to recognize and utilize psychological principles in the classroom as well as a more "traditional" paper and pencil exam would?"

The following results were obtained:

Yes 40 No 7 Uncertain 5

Students who responded favourable to the test also made the following remarks:

"Students are able to react to SOMETHING DIFFERENT. The tapes are far from boring, and there is relief from mad furious writing".

"The test makes the questions more realistic".

"Silence is much better portrayed on the screen than on paper".

"We could see how members responded to one another physically, in addition to their verbal remarks".

"They are terrific for pointing out certain areas and they relieve tensions on the part of the candidates".

"You have to know what your talking about to be able to recognize a process and write about it simultaneously".

"real life situation gives students an opportunity to apply anything they have learned. This is a TV generation. They are used to watching and learning from TV".

"The situation is visible and more immediate. Therefore easier to transfer to new situations".

On the negative side the following comments were also noted:

"technical problems in audio and visual recording and transmitting".

"not enough time to think out answers".

"difficulty to see and hear".

"The strain to cat everything cuases person to miss important parts".

"I can't concentrate while the film is running. Possibly you could leave a longer time span between showings to think about the first showing".

"necessity to think on the spot and make a decision..... need more experience before you can become adept at this; also whether you are right or wrong adds to the difficulty".

Final Remarks

The raison d'etre of this paper is to provide some evidence that videotape tests provide some alternative to more traditional tests. It is the author's contention that they enable the educator to assess student performance in a situation which is at least two steps closer to the realities of classroom life than is normally allowed for in regular paper and pencil tests.

The use and development of these tests does have some drawbacks. Media center specialists exert pressures to make the tests into major productions. Technical failures in production and showing can occur. Arrangements for classrooms and playback equipment need to be done well in advance of the testing date. (Even then such arrangements are made well in advance there is some risk that the equipment requisition will be misplaced or ignored!) Considerable difficulty is experienced if the examiner is snowbound or sick.

These problems recognized, the use of videotapes for testing purposes seems to hold potential for assessing achievement of intended learning outcomes in a variety of subjects and at various levels of the cognitive domain. For instance, Science methods instructors should be able to make various tapes which will assess how well their students' acquired skills transfer to new laboratory situations. Surely, it is important for teachers of English or Communications courses to be able to recognize differences in the effect of a communication caused by variations in speech patterns

and non-verbal cues. Use of these tests in the areas of Child Psychology, Abnormal Psychology, Classroom Psychology, Biology, Physical Science, topological problems in Mathematics, Social Studies and so on, seem quite obvious and one can only wonder why more energy has not been devoted to their development.

Critics are welcome to claim that such tests only represent another way of asking questions. But maybe that's what is needed in Education from time to time.